the plant journal

SEB Society for Experimental Biology

The Plant Journal (2018) **94,** 274–287 doi: 10.1111/tpj.13853

Organization of the TC and TE cellular T-DNA regions in *Nicotiana otophora* and functional analysis of three diverged TE-*6b* genes

Ke Chen[1,2], François Dorlhac de Borne[3], Nicolas Sierro[4], Nikolai V. Ivanov[4], Malek Alouia[1], Sandrine Koechler[1] and Léon Otten[1,*]

[1]*Department of Molecular Mechanisms of Phenotypic Plasticity, Institut de Biologie Moléculaire des Plantes, Rue du Général Zimmer 12, 67084, Strasbourg, France,*

[2]*Key Laboratory of Systems Biomedicine (Ministry of Education), Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China,*

[3]*Imperial Tobacco Bergerac, La Tour, 24100, Bergerac, France, and*

[4]*PMI R&D, Philip Morris Products S.A. [part of Philip Morris International group of companies], Quai Jeanrenaud 5, 2000, Neuchâtel, Switzerland*

Received 10 November 2017; revised 17 January 2018; accepted 22 January 2018; published online 3 February 2018.
*For correspondance (e-mail leon.otten@ibmp-cnrs.unistra.fr).

SUMMARY

***Nicotiana otophora* contains *Agrobacterium*-derived T-DNA sequences introduced by horizontal gene transfer (Chen *et al.*, 2014). Sixty-nine contigs were assembled into four different cellular T-DNAs (cT-DNAs) totalling 83 kb. TC and TE result from two successive transformation events, each followed by duplication, yielding two TC and two TE inserts. TC is also found in other *Nicotiana* species, whereas TE is unique to *N. otophora*. Both cT-DNA regions are partially duplicated inverted repeats. Analysis of the cT-DNA divergence patterns allowed reconstruction of the evolution of the TC and TE regions. TC and TE carry 10 intact open reading frames. Three of these are TE-*6b* genes, derived from a single *6b* gene carried by the *Agrobacterium* strain which inserted TE in the *N. otophora* ancestor. *6b* genes have so far only been found in *Agrobacterium tumefaciens* or *Agrobacterium vitis* T-DNAs and strongly modify plant growth (Chen and Otten, 2016). The TE-*6b* genes were expressed in *Nicotiana tabacum* under the constitutive 2 × 35S promoter. TE-1-*6b-R* and TE-2-*6b* led to shorter plants, dark-green leaves, a strong increase in leaf vein development and modified petiole wings. TE-1-*6b-L* expression led to a similar phenotype, but in addition leaves show outgrowths at the margins, flowers were modified and plants became viviparous, i.e. embryos germinated in the capsules at an early stage of their development. Embryos could be rescued by culture *in vitro*. The TE-*6b* phenotypes are very different from the earlier described *6b* phenotypes and could provide new insight into the mode of action of the *6b* genes.**

Keywords: *Nicotiana otophora*, cT-DNA, natural transformant, *6b* oncogene, *plast* genes, *Agrobacterium rhizogenes*, vivipary.

INTRODUCTION

Several plant species carry *Agrobacterium* T-DNA sequences (cellular T-DNAs or cT-DNAs) in their genomes and can be considered as 'natural transformants'. cT-DNAs were most likely introduced by infection with *Agrobacterium rhizogenes* and subsequent regeneration of hairy roots (Matveeva and Lutova, 2014; Chen and Otten, 2017). cT-DNAs have been found in *Nicotiana* (White *et al.*, 1983; Furner *et al.*, 1986; Chen *et al.*, 2014), *Linaria vulgaris* (Matveeva *et al.*, 2012) and *Ipomoea batatas* (Kyndt *et al.*, 2015).

Expression of the *Nicotiana glauca* genes Ng*orf13* and Ng*rolC* and of the *Nicotiana tabacum* genes *torf13* and *trolC* in various test plants demonstrated that they have growth-modifying properties similar to their *A. rhizogenes* equivalents (Meyer *et al.*, 1995; Fründt *et al.*, 1998a,b; Aoki and Syono, 1999b,c, 2000; Aoki, 2004; Mohajjel-Shoja, 2010; Mohajjel-Shoja *et al.*, 2011). The TB-*mas2'* gene of *N. tabacum* encodes opine synthesis in tobacco (Chen *et al.*, 2016).

The cT-DNA sequences have been identified and fully assembled using genomic sequence data. Four cT-DNA

inserts (TA, TB, TC and TD) were found in *Nicotiana tomentosiformis* (Chen *et al.*, 2014; Chen, 2016); these differ from the *N. glauca* cT-DNA. TA, TB, TC and TD carry partial, inverted repeats. Using repeat divergence, the order of the insertion events could be reconstructed as TC > TB > TD > TA. Members of the section Tomentosae carry different cT-DNA combinations (Chen *et al.*, 2014). From this analysis it was concluded that the *N. tomentosiformis* ancestors were repeatedly transformed by different *A. rhizogenes*-like strains. Lack of T-DNA-like sequences in the *Nicotiana setchellii* transcriptome (Long *et al.*, 2016), and the failure to amplify TC fragments from its genomic DNA (Chen, 2016), indicate that the *N. setchellii* line split off before the first cT-DNA insertion. The proposed order and distribution of the four cT-DNA inserts is consistent with the phylogeny of section Tomentosae (Knapp *et al.*, 2004).

Here we report on the cT-DNA sequences in *Nicotiana otophora* Griseb. This species was first described by August Grisebach from a population in Tarija, Cuesta Colorado, South Bolivia (Grisebach, 1879). The name *otophora* means 'having auricles'» and refers to the shape of the leaf petioles. Several *N. otophora* accessions have been collected from Bolivia and Argentina. The majority of the Tomentosae species are day-flowering and pollinated by bees and hummingbirds, but *N. otophora* is exceptional as it is night-flowering and pollinated by bats and hawkmoths (Nattero *et al.*, 2003). Nectar sugars and amino acids play a role in pollinator choice (Tiedge and Lohaus, 2017). *Nicotiana otophora* was initially suspected to be the paternal parent of *N. tabacum* (Goodspeed, 1954), but this idea was later abandoned, favouring instead *N. tomentosiformis* (Nattero *et al.*, 2003). *Nicotiana otophora* has been used in tobacco breeding (Reed and Schneider, 1992).

In a preliminary analysis, genomic sequences from *N. otophora* (Sierro *et al.*, 2014) were searched for cT-DNA sequences. This yielded 25 contigs with TC sequences and 44 contigs with other cT-DNA sequences. The latter were tentatively attributed to a new cT-DNA called TE (Chen *et al.*, 2014). The TC and TE sequences could not be assembled at the time because of unexpected sequence variability, suggesting three or four non-identical copies.

Surprisingly, the TE contigs carried not only *A. rhizogenes* sequences but also genes typical for *Agrobacterium tumefaciens*/*Agrobacterium vitis*: vitopine synthase (*vis*) and gene *6b*. The *6b* gene was initially described as an oncogene since it induces tumours on *N. glauca* (Hooykaas *et al.*, 1988). Its expression in various plant species causes strong morphological changes (for reviews see Ishibashi *et al.*, 2014; Ito and Machida, 2015). AB-*6b* from *A. vitis* strain AB4, and T-*6b* from *A. vitis* strain Tm4 have been expressed in tobacco under 2 × 35S promoter control, and induce a specific phenotype, called 'enation syndrome' (Helfer *et al.*, 2003; Grémillon *et al.*, 2004). This includes leaf and flower doubling (enations and catacorollas, respectively), ectopic vascular strands, root thickening on sucrose media and a large number of other abnormalities. Four basic *6b*-induced changes were identified: chlorosis, induction of ectopic vascular strands and leaf primordia, and abnormal cell expansion (Chen and Otten, 2016). Earlier work showed a *6b*-induced increase in sucrose uptake and retention, both in leaf discs and root fragments, suggesting that this could be the underlying mechanism for the observed growth changes (Clément *et al.*, 2006, 2007). Other mechanisms have been proposed for *6b* gene activity (Wabiko and Minemura, 1996; Gális *et al.*, 2002, 2004; Kitakura *et al.*, 2002, 2008; Kakiuchi *et al.*, 2006; Terakura *et al.*, 2007; Wang *et al.*, 2011; Takahashi *et al.*, 2013; Ishibashi *et al.*, 2014; Ito and Machida, 2015; summarized in Chen, 2016).

The unexpected finding of *vis* and *6b* sequences in *N. otophora* raised the following questions. Are they connected to the *A. rhizogenes*-like sequences or do they result from a transformation by *A. tumefaciens* or *A. vitis*? Do the cT-DNA sequences of *N. otophora* contain intact open reading frames? Are the unusual *6b* genes intact and, if so, are they biologically active? Preliminary analysis of the contigs showed that TC and TE were duplicated, partial inverted repeats, yielding from one to four copies for different regions. This complex situation resulted in errors in the automatic contig construction and necessitated reassembly and re-mapping of the original reads. Here we report the maps of the TC and TE inserts, a reconstruction of the evolution of these inserts, and a functional analysis of the *N. otophora* TE-*6b* genes.

RESULTS

Properties of *N. otophora*

We obtained six *N. otophora* accessions: TW94, TW95, TW96, TW97 (from the US Nicotiana Germplasm Collection), NIC406 (from IPK Gatersleben) and ITB643 (from Imperial Tobacco Bergerac). These accessions show phenotypic differences, particularly in leaf venation (Figure S1a–f in the online Supporting Information; Figure S1g shows *N. tabacum* cv. Samsun). Accessions with well-developed veins have wrinkled leaves (Figure S1h), petiole wings are large and carry strong veins like the leaf (Figure S1i). These differences were stable and reproducible under our greenhouse conditions. Since *N. otophora* TW95 was used to obtain genomic and transcriptome sequence data (Sierro *et al.*, 2014), we used this accession for our studies.

Mapping of TC-1 and TC-2

The *N. tomentosiformis* TC region (TOF-TC) has been described before (Chen *et al.*, 2014). It is a 17-kb imperfect inverted repeat of a DNA fragment which resembles the

left part of the *A. rhizogenes* A4 TL-DNA (Slightom *et al.*, 1986) with *orf2*, *orf3n*, *orf8*, *rolA* and *rolB* sequences. The inverted repeat is flanked by two unique regions: an *ocl* (octopine synthase-like) gene in the left part of the cT-DNA insert and a gene *c*-like sequence on the right. TOF-TC carries only one intact open reading frame (ORF), *ocl*. The *ocl* and *c*-like genes are unusual, as they are normally only found in *A. tumefaciens* or *A. vitis*.

Nicotiana otophora TW95 carries two TC copies, TC-1 and TC-2 (Figure 1). These differ by 4%, but are inserted in the same plant DNA sequence as TOF-TC and are similarly organized, therefore they are derived from the same insertion event. The *N. otophora* and *N. tomentosiformis* TC regions show some interesting differences. TC-1 carries a 3237-nucleotide (nt) plant sequence in the right arm of its inverted repeat, within the *orf3n* sequence (TC-1 plant sequence; TC-1P, Figure 1a), surrounded by a 16-nt direct repeat (TATCATTCTCGCATCA). TC-1P is 72% identical to a long interspersed nuclear element (LINE-1-like retrotransposon) from *Solanum tuberosum* (HM013964.1; Wolters *et al.*, 2010). TC-1P potentially encodes a 782-amino-acid ORF with 76% identity to an RNA-directed DNA polymerase from mobile element Jockey-like from *Nicotiana attenuata* (XP_019236372.1). TC-1P sequences are also found elsewhere in the *N. otophora* genome. TC-2 carries a 206-nt plant sequence in the left arm (TC-2P, Figure 1a), surrounded by an imperfect 12-nt direct repeat (TTGTGCAAACTA and TTGTCAAAACTA), 73 nt downstream of the TC-1-*orf3n* stop codon. This sequence is similar to various *Nicotiana* short interspersed nuclear element (SINE) sequences (Wenke *et al.*, 2011), and is partially present in TC-1P. TC-2P is 95% identical to SINE TS_Nt2 (221 nt, HE583509) from *N. tabacum* and is also found elsewhere in *N. otophora*. LINES and SINES are reverse transcripts of RNA molecules, integrated in various genomic locations (Wenke *et al.*, 2011). TOF-TC lacks TC-1P and TC-2P. Therefore it is likely that these sequences were inserted into the *N. otophora* TC regions after the separation of *N. otophora* and *N. tomentosiformis*.

The structure of the *orf3n–orf8* regions of the right and left arms of the TC-1 and TC-2 regions of *N. otophora* and *N. tomentosiformis* carry not only TC-1P and TC-2P but also other indels. Together, they are marked 'a' to 'h' in the sequence alignment shown in Figure S2 and in the schematic in Figure 1(b). The distribution of the various indels was used to reconstruct the evolution of this region (Figure 1b). A 35-nt deletion ('b') is present in all right arms, and therefore occurred at an early stage after insertion, before TC duplication. A 350-nt deletion in the left arm of TC-1 of both *N. otophora* and *N. tomentosiformis* ('h') must have occurred prior to the separation of the *N. tomentosiformis* and *N. otophora* lines. A 17-nt deletion ('c') and the TC-1P insertion ('a') in the right arm of TC-1 of *N. otophora,* but not of *N. tomentosiformis*, must have occurred after the separation of *N. tomentosiformis* and *N. otophora*. Finally, a 9-nt insertion ('g') and the TC-2P insertion ('e') in the TC-2 left arm, and an 8-nt deletion ('d') and 31-nt insertion ('f') in the TC-2 right arm are specific for TC-2 and therefore occurred after the TC duplication.

The different TC modifications provide a scenario for the evolution of the TC region (Figure 1b, c). Assembly of the two *N. otophora* TC regions allowed us to identify their intact ORFs. TC-1 carries two intact ORFs, *ocl* and *orf3n*, whereas TC-2 contains none.

Mapping of TE-1 and TE-2

The TE regions were assembled in the same way as were the TC regions. As with TC, two TE copies were found: TE-1 and TE-2 (Figure 2).

The common parts of TE-1 and TE-2 differ by 1.7%, with TE-1 being larger than TE-2. On the left of Figure 2, TE-1 resembles the T1 T-DNA of *A. vitis* strain S4 (Canaday *et al.*, 1992). However, similarity between TE and S4 T-1 is low, and only the *vis* part was detected by BLASTN analysis (71% nucleotide identity with region 3122–2084 of M91608.1). The closest TE-*6b* homolog is T-*6b* from *A. vitis* Tm4 (71% nucleotide identity, but only for the central 174–504 part of the *6b* gene, on a total of 630 nt). The right end of the TE-1 *vis-6b* region is part of a partial repeat (R1) with *6b* and tryptophan monooxygenase (*iaaM*) sequences. The left and right arms (R1-L and R1-R) are separated by a unique region with *iaaH* and agrocinopine synthase (*acs*) sequences. The organization of R1 is similar to that of the *A. tumefaciens* and *A. vitis* T-DNAs of the LB-*acs-5-iaaH-iaaM-ipt-6a-6b-ocs/vis*-RB type (as in *A. tumefaciens* A6, or *A. vitis* Tm4). However, genes *5*, *ipt* and *6a* are missing. The right part of TE-1 consists of another inverted repeat (R2) with *rolC-orf13-orf13a-orf14-mas2′-mas1′* sequences on its left and right arms (R2-L and R2-R). R2 is similar in organization and sequence (78% identity) to the right part of the *A. rhizogenes* strain 8196 T-DNA (Hansen *et al.*, 1991).

TE-2 is similar to TE-1, but the right arm of R1-R and the left arm of R2-L are missing. The region between R1-L and R1-R (with *iaaH* and *acs* sequences) is inverted with respect to TE-1, and with respect to the normal T-DNA orientation. This is probably due to homologous recombination between the R1 repeats.

A model for the origin and evolution of the TE inserts is shown in Figure 3. The partial inverted repeats were most likely derived from two separate T-DNAs (marked in blue and red): one with *acs-iaaH-iaaM-6b-vis* and the other with *rolB-rolC-orf13-orf13a-orf14-mas2′-mas1′*, linked at their right borders, and (most likely) located on the same plasmid. Alternatively, TE could result from a mixed infection involving two different *Agrobacterium* strains. Fragments a, b, c and d were probably ligated together before integration to form a composite structure, which was then inserted into the plant genome. Subsequently, this insert

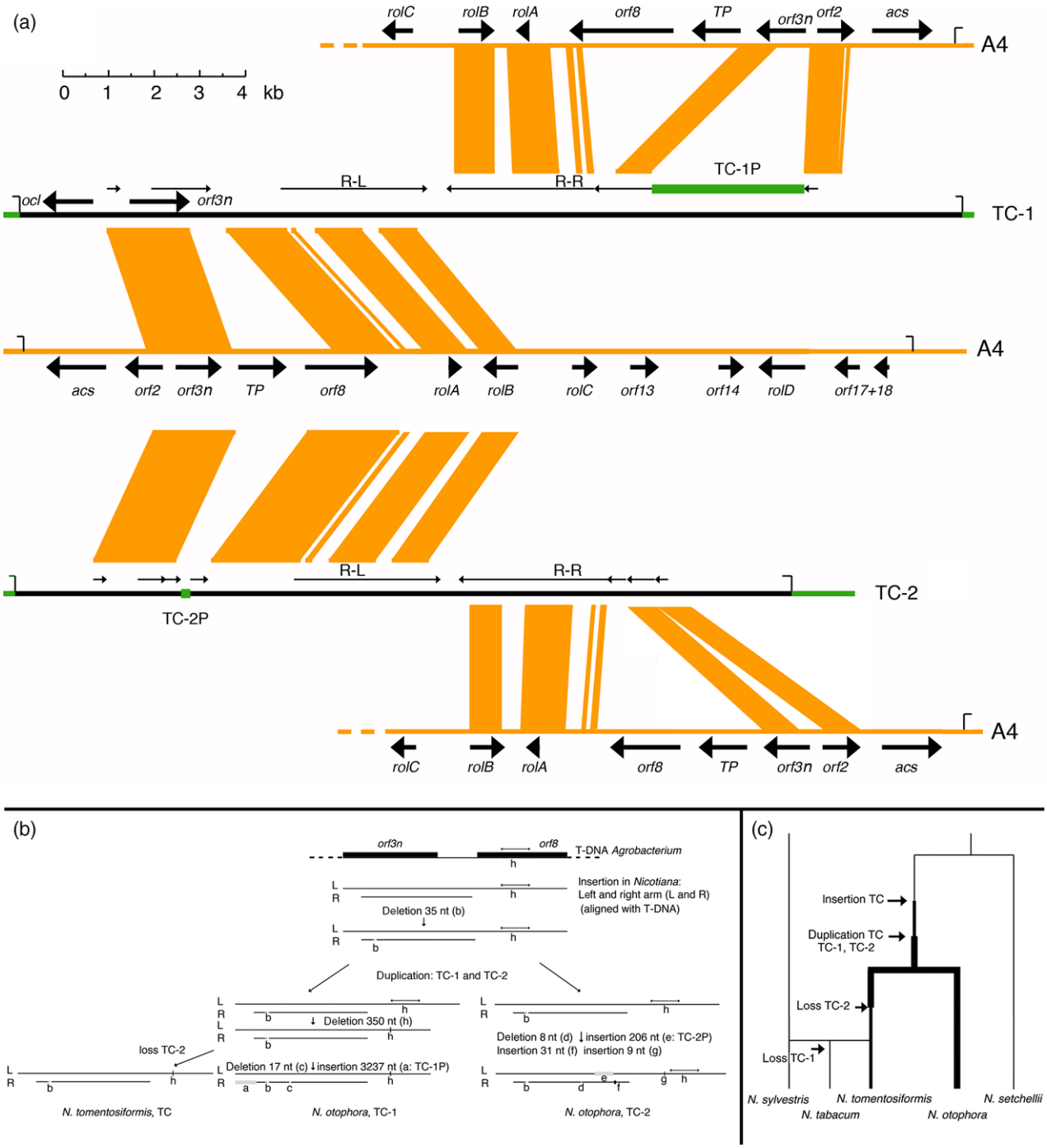

Figure 1. TC-1 and TC-2.

(a) Maps of TC-1 and TC-2. Horizontal black lines: TC regions. Horizontal orange lines: *Agrobacterium rhizogenes* A4 TL-DNA (Slightom *et al.*, 1986), with open reading frames (ORFs). Orange blocks: regions of similarity with A4 TL-DNA. Only the intact TC ORFs (TC-*ocl* and TC-*orf3n*) are shown. R-L and R-R: left and right arms, respectively, of inverted repeats. Green areas: plant DNA. TC-1P and TC-2P: plant insertion elements.

(b) Evolution of the TC *orf3n-orf8* region. This model is derived from indel events (a–h) found in the *Nicotiana tomentosiformis* TC region (Chen *et al.*, 2014) and in TC-1L, TC-1R, TC-2L and TC-2R from *Nicotiana otophora* (this work). For sequence alignment see Figure S2. Left and right arms are aligned in the same direction to facilitate comparison. Evolution of regions *orf3n*–*orf8* is presented from top to bottom. The 350-nucleotide (nt) sequence marked 'h' was deleted after the TC duplication event in TC-1 but not in TC-2. *Nicotiana tomentosiformis* inherited TC-1. TC-1P (a) and TC-2P (e) are plant-derived insertions.

(c) Evolution of the TC region. The width of lines indicates zero, one and two copies. After the line leading to *Nicotiana setchellii* had split off, a single TC region was inserted. This TC region was subsequently duplicated to TC-1 and TC-2. TC-2 was lost in the line leading to *N. tomentosiformis*. In *Nicotiana tabacum*, TC-1 was lost as well.

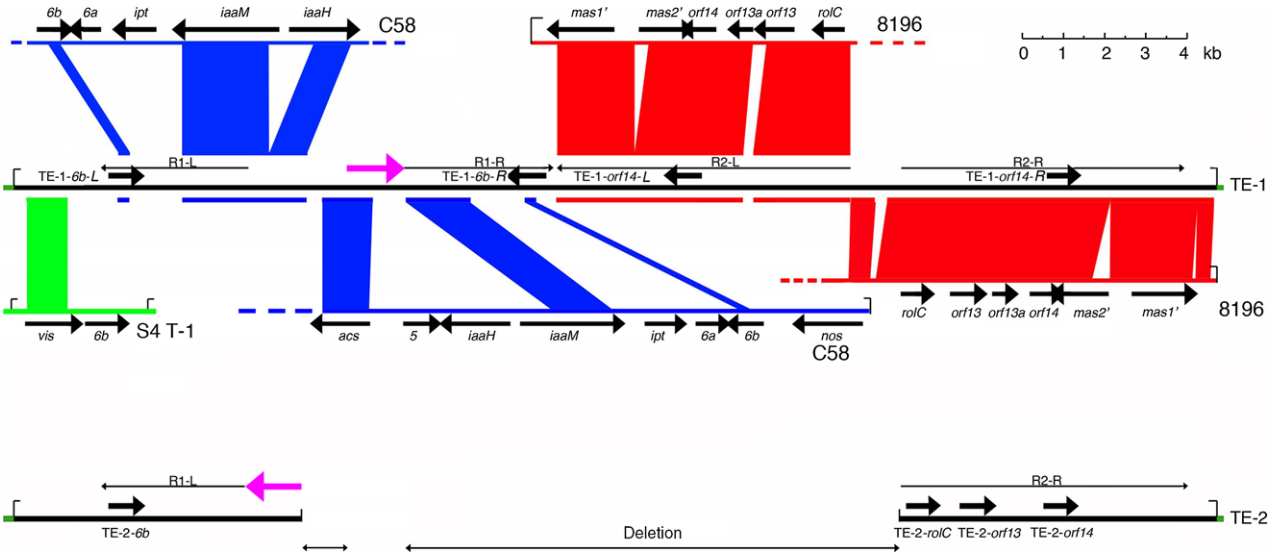

Figure 2. Maps of TE-1 and TE-2.
Horizontal black lines: TE regions. Green lines: *Agrobacterium vitis* S4 T-1 DNA with *6b* and *vis* genes (Canaday *et al.*, 1992). Blue lines: *Agrobacterium tumefaciens* C58 T-DNA (AJ237588.1, right part). Red lines: right part of *Agrobacterium rhizogenes* 8196 T-DNA (Hansen *et al.*, 1991). Coloured blocks indicate DNA similarity. Only intact TE-1 and TE-2 ORFs are shown. R1-L, R1-R, R2-L and R2-R: left and right arms, respectively, of R1 and R2 repeats. Magenta arrow: unique region with different orientations in TE-1 and TE-2, the orientation in TE-1 is the normal one. In TE-2, the central region of TE-1 (R1-R and R2-L) is deleted.

was duplicated and both regions underwent further modifications.

Assembly of the two TE maps allowed us to identify eight intact ORFs: TE-1-*6b-L*, TE-1-*6b-R*, TE-1-*orf14-L*, TE-1-*orf14-R*, TE-2-*6b*, TE-2-*rolC*, TE-2-*orf13* and TE-2-*orf14*. All belong to the *plast* gene family (*plast* for phenotypic plasticity; Levesque *et al.*, 1988; Helfer *et al.*, 2002). The TE-6B and TE-Orf14 proteins are quite different from their homologues (see below) but TE-Orf13 and TE-RolC are very similar to the *N. glauca* equivalents (Figure S3).

Interestingly, TC carries *orf2-orf3-orf8-rolA-rolB* sequences (as on the left part of the A4 TL-DNA) whereas TE carries *rolC-orf13-orf13a-orf14-mas2′-mas1′* sequences (as on the right part of the A4 TL-DNA). Thus it seemed possible that both represent two different fragments from the same T-DNA. However, TC and TE show a small 350-nt overlap at the start of the *rolB* gene with only 77% identity. Therefore, TC and TE are derived from two different *A. rhizogenes* strains and were introduced at different times. In total, *N. otophora* contains 83 kb of cT-DNA, distributed over four inserts. No border sequences could be found (see Experimental Procedures). Together, TC-1, TC-2, TE-1 and TE-2 carry 10 intact ORFs.

Transcription of the TC and TE regions

The cT-DNA sequences of *N. otophora* were investigated for transcriptional activity (see Experimental Procedures). Transcriptome reads from TW95 leaf and root tissues were mapped on TC-1, TC-2, TE-1 and TE-2 (Figure S4). Overall transcription levels were low, except for TC-1P [13.9 and 6.4 RPKM (reads per kilobase of transcript per million mapped reads) in leaf and root, respectively] and TC-2P (3.4 and 6.8 RPKM in leaf and root, respectively). TC-1 and TC-2 had very low transcription levels (0.01 RPKM for TC-1-*ocl* in leaf), whereas TE-2 had higher levels than TE-1 (0.41 RPKM for TE-1-*6b-L* in root, 0.59 and 0.41 RPKM for TE-2-*6b* in leaf and root respectively, 0.44 RPKM for TE-2-*rolC* in leaf, 0.22 RPKM for TE-2-*orf13* in leaf, and 0.20 and 0.23 RPKM for TE-2-*orf14* in leaf and root, respectively).

Functional analysis of the *N. otophora 6b* genes

Nicotiana otophora TE-1 and TE-2 contain three intact *6b* genes, derived from the single *6b* gene carried by the original T-DNA (Figure 3). TE-1-*6b-L* is part of the TE-1 R1-L repeat and TE-2-*6b* occupies an equivalent position on TE-2. TE-1-*6b-R* is located on the R1-R part of TE-1; this region is deleted in TE-2. The three predicted 6B protein sequences are very similar to each other, and quite different from the closest homolog, T-6B (CAA39648.1) with an identity value of only 54% (113 out of 208 amino acid residues for TE-1-6B-L). The acidic repeat found in the C-terminal part of various 6B proteins (Helfer *et al.*, 2002) is also present in the TE-6B proteins (Figure S5). However, TE-1-6B-L is shorter by seven E residues. The promoter regions of TE-1-*6b-L* and TE-2-*6b* are similar. In contrast, 167 nt upstream from the start codon, the TE-1-*6b-R* promoter region is linked to the 3′ region of the *mas1′* gene on the R2-L repeat, and its sequence diverges. The transcription data show very low expression for TE-1-*6b-R*, and measurable levels for TE-1-*6b-L* and TE-2-*6b*, mainly in the leaves (Figure S4).

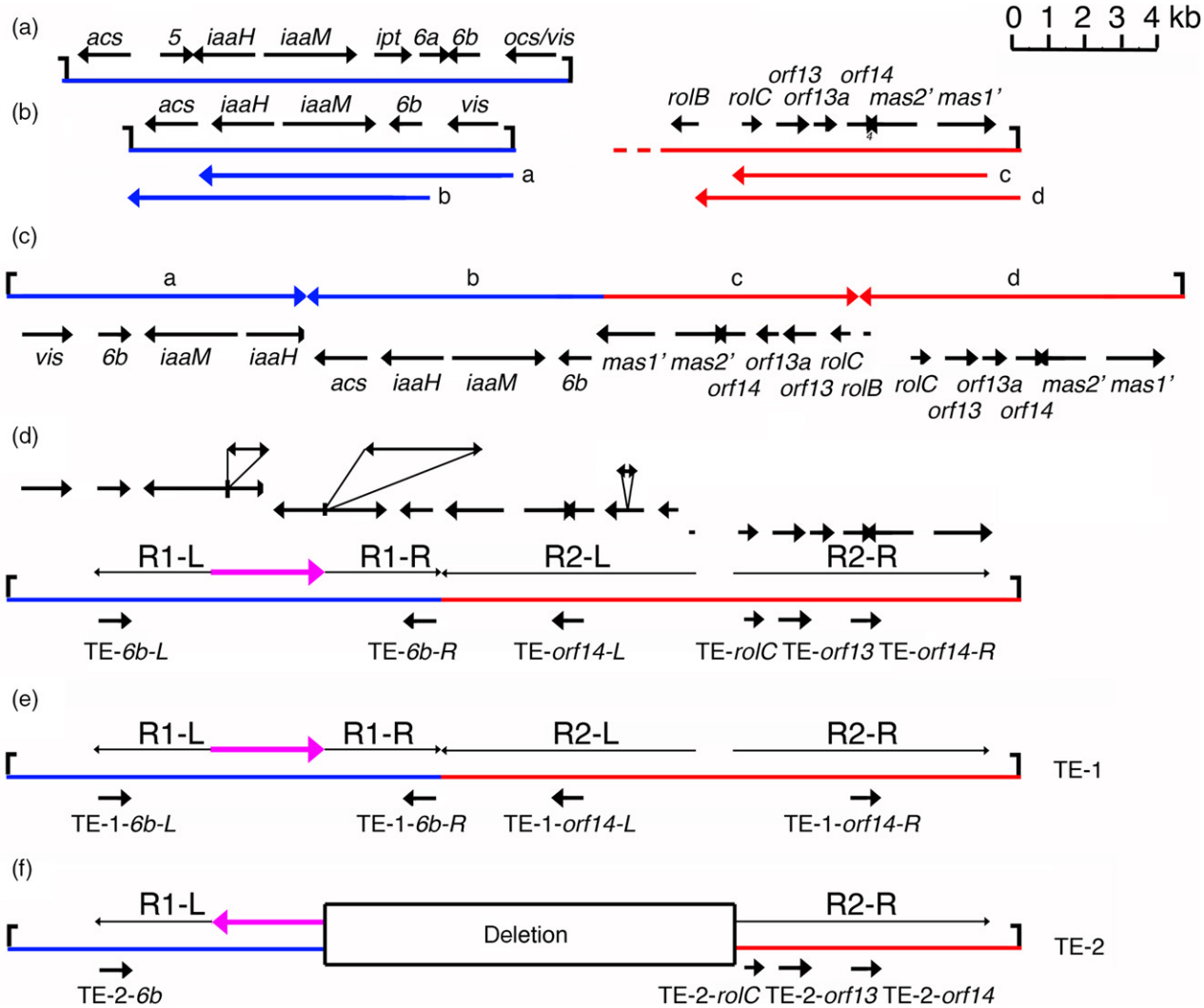

Figure 3. Model for the evolutionary origin of TE-1 and TE-2.
(a) Octopine T-DNA structure. (b) Hypothetical *Agrobacterium tumefaciens*-like T-DNA (blue) and *Agrobacterium rhizogenes*-like T-DNA (red) of the *Agrobacterium* strain which introduced TE. Most likely the blue and red regions were part of separate T-DNAs from the same Ri plasmid. On the left: an *acs-iaaH-iaaM-6b-vis* region, related to the octopine T-DNA structure *acs-5-iaaH-iaaM-ipt-6a-6b-ocs/vis*, but without genes *5, ipt* and *6a*. Fragments a and b became part of the TE region. On the right: right part of an *A. rhizogenes*-type T-DNA with genes *rolB-rolC-orf13-orf13a-orf14-mas2'-mas1'* (as in *A. rhizogenes* strain 8196; Hansen *et al.*, 1991). Fragments c and d became part of the TE region. (c) Original TE region obtained by ligation of T-DNA fragments a, b, c and d, shown in (b). It is assumed that originally all open reading frames (ORFs) were intact, except at the junctions between the fragments. (d) Modification of the original TE region by deletions (double-headed arrows), and loss of several ORFs. TE-*6b-L*, TE-*6b-R*, TE-*orf14-L*, TE-*rolC*, TE-*orf13* and TE-*orf14-R* remained intact. (e) Generation of TE-1 by duplication of the TE region, and loss of TE-1-*rolC* and TE-1-*orf13*. (f) Generation of TE-2 by duplication of the TE region, followed by loss of the central part with R1-R and R2-L, and inversion of the region indicated by the magenta arrow.

The three TE-*6b* genes were placed under the control of the constitutive 2 × 35S promoter (see Experimental Procedures) in order to maximize chances to obtain a phenotype, and tested for biological activity. As several *6b* genes induce tumours on *Nicotiana rustica* and *Kalanchoe tubiflora* (Hooykaas *et al.*, 1988; Helfer *et al.*, 2002; Chen and Otten, 2016), we first tested the TE-*6b* gene constructs on these plants, but no tumours were formed. Subsequently, the TE-*6b* genes were introduced into *N. tabacum* (see Experimental Procedures).

TE-1-6b-R and TE-2-6b primary regenerants (40 each) showed similar abnormalities. Three homozygous single-locus lines grew slowly and had wrinkled leaves (lines 5, 14 and 28; Figure 4a); RT-qPCR analysis showed relatively high TE-1-*6b-R* expression levels in these lines (Figure 4b). They also had more prominent tertiary and higher-order leaf veins than normal tobacco (Figure 4c), and their petiole wings formed leaflets with many large glandular trichomes (Figure 4d, e).

TE-1-6b-L primary regenerants (56 plants) were more abnormal than TE-1-6b-R and TE-2-6b regenerants. In

severe cases (Figure 5a–f), fan-shaped leaves developed. In some cases, leafy filaments developed on the leaf edge and grew into fan-shaped structures (Figure 5a, b). Leaf margins were broader and irregular and carried numerous large trichomes (Figure 5c, d). Roots grew directly from the leaf base (Figure 5e). Ectopic meristems could be found on the adaxial leaf surface (Figure 5f). Less severely affected, but still highly modified, TE-1-6b-L plants (Figures 6a and S6) developed stems and roots, and although they grew slowly they reached the same height as wild-type tobacco before flowering. Petiole wings formed distinct leaflets (Figure 5b). Veins were strongly developed, up to the leaf edge (Figure 6c–e). Leaf tips of TE-1-6b-L plants were often split (Figure 6f), and leaf edges could be highly irregular (Figure 6g). Leaflets growing from the leaf edges showed white, semiglobular growing points (Figure 6h, i), as on ectopic meristems in dex-T-*6b* tobacco plants (Chen and Otten, 2016). Ectopic shoot primordia appeared on some leaves (Figure 6j, k). The youngest leaves (Figure S6k) and the base of older leaves (Figure S6c) curled downward at the edges, and bulged between the veins (Figure S6d). Leaf edges on the apical part of the lower leaves were dentate at the apex and folded at the base (Figure S6a, b), and leaflets seemed to grow out from veins growing perpendicular to the leaf edge (Figures 6e–h and S6e–g, i), forming dentate leaves. Dentate leaves are found in numerous plants (such as *Lactuca virosa*; Figure S6h). Young leaves carried numerous large trichomes, giving them a tomentose (woolly) appearance (Figure S6j, k). On some leaves, spikes emerged at junctions of secondary veins (Figure S6l–n). RT-qPCR analysis of TE-1-*6b-L* expression in F_1 plants from the single-locus lines S34, S39, S42, S47 and S56 showed that strong phenotypes were correlated with high expression levels (Figure S7a, b). Expression of T-*6b* leads to sucrose uptake and accumulation (Clément *et al.*, 2006, 2007), causing dramatic root cell expansion in tobacco (Grémillon *et al.*, 2004; Clément *et al.*, 2007; Pasternak *et al.*, 2017). This is considered to be an important characteristic of *6b* activity (Chen and Otten, 2016). Homozygous S34 seedlings were therefore grown on M0222 medium with 2% sucrose. S34 seedlings (Figure S7d–f) grew more slowly than the controls (Figure S7c) and had abnormal cotyledons. However, S34 roots were indistinguishable from control roots (Figure S7g–j). Thus, TE-1-*6b-L* does not induce root expansion on 2% sucrose, unlike T-*6b*.

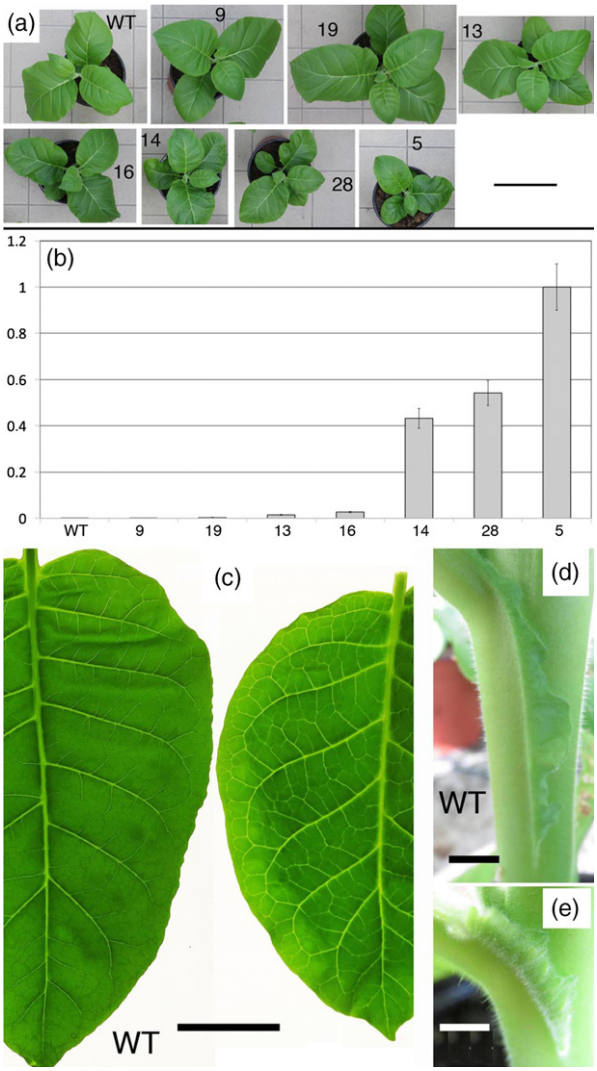

Figure 4. Phenotypes of TE-1-6b-R transformants.
(a) Wild type tobacco (WT) and seven homozygous single-locus transformants. Lines 5, 14 and 28 are smaller.
(b) RT-qPCR for TE-1-*6b*-R for the lines shown in (a). Values were calculated relative to the highest value (line 5). The stunted lines 5, 14 and 28 have a high TE-1-*6b-R* expression level.
(c) Leaf veins in WT tobacco (left) and TE-1-6b-R line 5 (right). The transformed line shows more prominent tertiary and higher-order veins.
(d) Petiole wings from normal tobacco.
(e) Petiole wings from TE-1-6b-R line 5: small leaflets with many trichomes replace the normal petiole wings.
Scale bars: (a) 20 cm; (c) 5 cm; (d), (e) 2 cm.

TE-1-*6b-L* induces vivipary

Most TE-1-6b-L plants produced flowers with an altered phenotype. Petals were less coloured and curved downwards (Figure 7a, b). Remarkably, most TE-1-6b-L lines were viviparous: their seeds germinated prematurely in the capsules, forming hundreds of embryo-like structures (Figures 7c–h and S8). Capsules from flowers shortly after pollination contained embryo-like structures at the heart stage (Figure S9c, e). These embryos seemed to rupture the seed integuments and grew rapidly in size, forming green tubes. As a consequence, the capsules became larger than normal (Figure 7e). Capsules of a few TE-1-6b-L lines (e.g. S39 and S34) contained both normal seeds and germinating embryos, grouped in patches (Figure S8c, d). At a later stage, capsules burst open and embryos formed long green hypocotyls with cotyledons (Figures S8e and S7k).

The embryos eventually turned black and dried out (Figure S8f). Capsules of single-locus TE-1-6b-L lines contained only germinated embryos, showing that premature seed germination is determined by the parental plant. Embryos from viviparous plants looked initially quite similar, but later became heterogeneous in shape (Figure 7h). Physical constraints inside the capsule probably contributed to unequal development. The prematurely germinating structures did not develop into normal plantlets but had a white base and one or two abnormal, dark-green cotyledons. The embryo apex emerged first from the seed coat (Figure S9c, d), contrary to the apex of wild-type seedlings, which emerges last (Figure 7i). No roots were formed (Figures 7j, k and S8–S10). Some embryos formed spherical, stalked structures in the capsules (Figure S9a, b). Heart-shaped early stage embryos could be grown in M0222 medium with 1% sucrose for 3 weeks (Figure S9e, f), and then became necrotic. Stalked structures like those seen in the capsules emerged and grew out into longer structures (Figure S10). Further development (Figure S10h) showed that they were leaf primordia. Thus, TE-1-6b-L embryos initiate shoot apical meristems and leaf primordia, but their development differs considerably from that of normal tobacco embryos. Some shoot meristems formed between the cotyledons (Figure S10a, b), others below the cotyledons on the hypocotyl (Figure S10c, d), or on one of the cotyledons (Figure S10e). Secondary shoot meristems appeared on some embryos (Figure S10b, g). On M0222 medium with 1% sucrose the embryos did not form roots. Instead, a mass of cell filaments grew at the base of the embryo (Figure S10a–d), possibly derived from early root hairs (Figure 7i).

DISCUSSION

Nicotiana otophora contains 83 kb of *Agrobacterium*-derived DNA, consisting of two different cT-DNA types, TC and TE. TC and TE each occur in two copies, due to duplication of the original inserts with their surrounding plant DNA.

TE was inserted after TC, as shown by the divergence of the inverted repeats (1.7% and 6%, respectively). Since the rate of nucleotide divergence in *Nicotiana* is about 6% per million years (Lim *et al.*, 2007), TC and TE are estimated to be 1 and 0.3 million years old, respectively. *Nicotiana tomentosiformis* has only one TC copy which corresponds to TC-1. The loss of a 350-nt T-DNA fragment in TC-1 but not in TC-2, and the inheritance of this TC-1 deletion by *N. tomentosiformis* show that the TC duplication occurred before the separation of the lines leading to *N. otophora* and *N. tomentosiformis*. As the TC-1/TC-2 divergence (4%) is greater than the TE-1/TE-2 divergence (2%), it is likely that TC was duplicated before TE. A better understanding of these duplication events may be obtained by determining the full extent of the duplicated regions, both in *N. otophora* and in other Tomentosae species.

Only two *N. otophora* TC genes are intact: TC-1-*ocl* and TC-1-*orf3n*. 35S-A4-*orf3n* causes growth abnormalities in tobacco (Lemcke and Schmülling, 1998). Thus, TC-1-*orf3n* should be tested for growth effects and compared with its A4 homolog. *Nicotiana tomentosiformis* TC-*ocl* is intact, but its overexpression in *N. benthamiana* failed to produce opines (Chen *et al.*, 2014). In *N. tabacum*, TC has been completely deleted (Chen *et al.*, 2014). The highly degenerate nature of the TC region and its absence in *N. tabacum* seem to indicate that most of the initial TC genes lacked a selective advantage. However, they could have played a role at an early stage, favouring regeneration or reproductive isolation, two of the steps postulated for the survival of natural transformants (Chen and Otten, 2017). TC-1 and TC-2 carry plant transposon-like sequences. TC-1P (3237 nt) is similar to a LINE sequence and is inserted in TC-1-*orf3n*, TC-2P (206 nt) corresponds to a SINE sequence and is inserted immediately 3′ of TC-2-*orf3n*. The *N.*

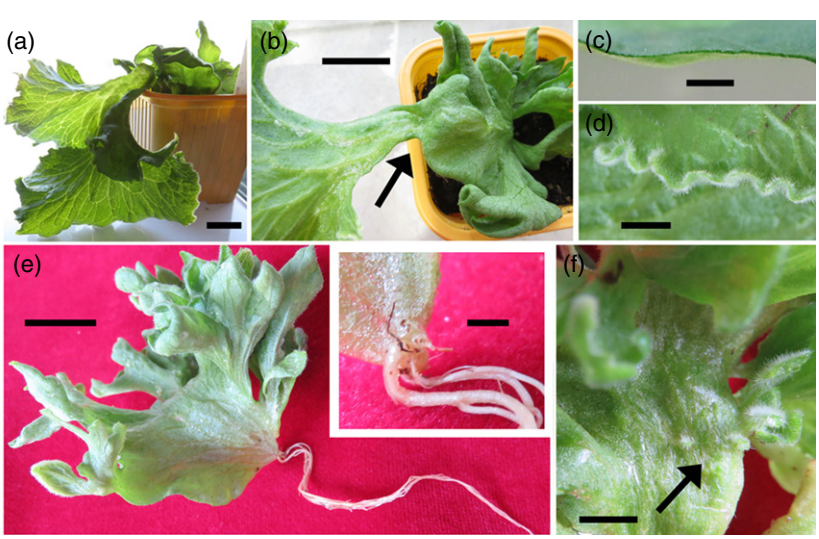

Figure 5. Highly abnormal TE-1-6b-L regenerants.
(a) Fan-shaped leaf-like structure with well-developed vascular system. (b) Detail of the structure in (a); the fan-shaped leaf is connected to another leaf-like structure by a ribbon-like structure (arrow) and appears as an outgrowth from the rim of the older structure. (c) Edge of a normal tobacco leaf. (d) Edge of the TE-1-6b-L leaf shown in (a) and (b). The edge of the TE-1-6b-L leaf is thicker, with many trichomes. (e) Rooting of another abnormal fan-shaped TE-1-6b-L leaf. Inset: roots emerge directly from the leaf. (f) Growth of buds (arrow) on leaf-like structures as in (a) and (d).
Scale bars: (a), (b), (e) 2 cm; (c), (d) 1 cm; (e) inset and (f) 2 mm.

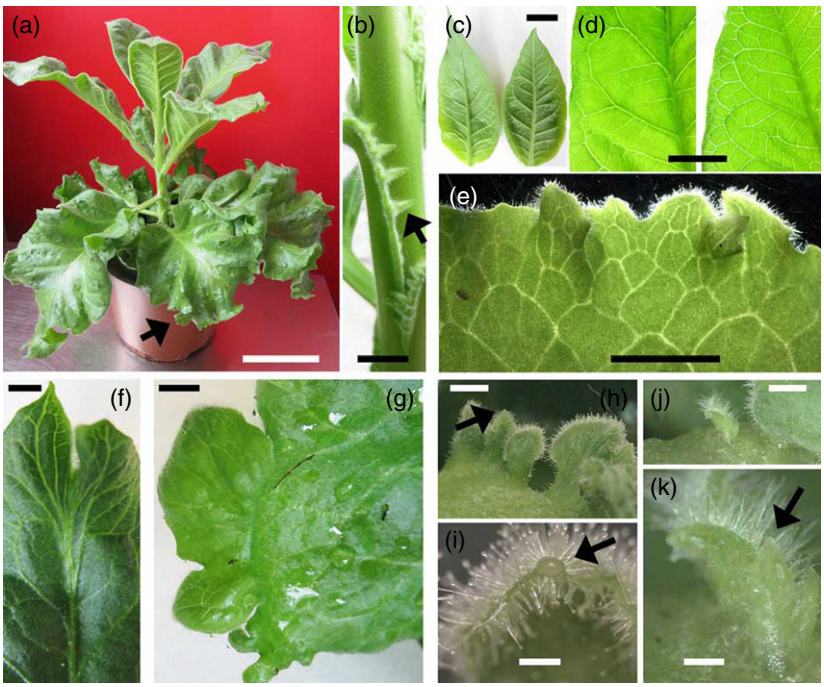

Figure 6. Abnormal TE-1-6b-L regenerants growing in soil.
(a) TE-1-6b-L tobacco plant in soil. The lower leaves are strongly wrinkled, folded down along the rim and dark-green. Upper leaves have downward-curving leaves. (b) Petiole wings of TE-1-6b-L regenerant with leaf-like structures. (c) Vein pattern in normal tobacco (left) and a TE-1-6b-L plant (right). (d) Detail of the vein pattern in normal tobacco (left) and a TE-1-6b-L plant (right). Veins of TE-1-6b-L plants are more prominent and extend up to the leaf edge. (e) Detail of TE-1-6b-L leaf. The leaf edge forms small leaf-like structures. (f) Leaf tip of a TE-1-6b-L plant: aberrant veins and splitting of the leaf apex. (g) Highly irregular leaf edge of a TE-1-6b-L plant. (h) Growth of leaflets on the leaf edge of a TE-1-6b-L plant. The arrow shows the leaflet tip. (i) Detail of leaflet tip shown in (h). (j) Small buds growing on a TE-1-6b-L plant. (k) As (j), detail of a bud tip.
Scale bars: (a) 15 cm; (b), (c) 2 cm; (d) 4 cm; (e)–(g) 1 cm; (h), (j) 4 mm; (i), (k) 1 mm.

tomentosiformis TC region (similar to TC-1) does not carry such elements, showing that TC-1P was inserted after TC duplication. Plant transposon-like sequences have also been found in a cT-DNA of *I. batatas*. The *Ib*T-DNA1 region of *I. batatas* variety 'Huachano'» carries a 6.8-kb Gypsy 2 type long terminal repeat transposon, but is absent in the variety 'Xu781' (Kyndt *et al.*, 2015).

TE-1 and TE-2 harbour eight intact genes with potential morphogenetic activity: three *6b* genes, three *orf14* genes, one *orf13* gene and one *rolC* gene. All belong to the *plast* gene family. Expression analysis showed low, but significant, expression for these ORFs, but also for degenerate genes interrupted by stop codons. Transcription of degenerate genes could be due to residual promoter activity.

Among the *N. otophora* cT-DNA genes, the three *6b* genes are especially interesting since they are linked to typical *A. rhizogenes* T-DNA genes. *Agrobacterium* plasmids with *acs-iaaH-iaaM-6b-vis* genes on one T-DNA and *rolB-rolC-orf13-orf13a-orf14-mas2′-mas1′* on another T-DNA have not been found so far. However, if such a plasmid survived, its sequences should still be similar to the ancestral plasmid, as indicated by the low level of divergence between the TE repeats. The transfer of the *acs-iaaH-iaaM-6b-vis* T-DNA from such a plasmid may have induced tumours and the transfer of the *rolB-rolC-orf13-orf13a-orf14-mas2′-mas1′* T-DNA may have induced roots. In the case of a combined transfer, one or the other effect could dominate, or intermediate structures could be formed, conferring a hybrid character on such a plasmid. It is therefore impossible to say at this point whether the plasmid carrying these T-DNAs was a Ti or Ri plasmid.

Reconstruction of the evolution of TE shows that the original TE insert contained two identical *6b* genes located on the R1 repeat. The insert was then duplicated, yielding four *6b* copies, but one of these, TC-2-*6b-R*, was lost by the deletion of the R1-R/R2-L region (Figure 3). During the evolution of TE the repeats and the *6b* copies diverged, both structurally and functionally. Expression in tobacco in the same expression cassette showed that TE-1-6B-L protein has a stronger activity than TE-1-6B-R and TE-2-6B. Site-directed mutagenesis could identify the amino acids responsible for the weak and strong phenotypes.

Unexpectedly, the TE-*6b* phenotypes were found to differ considerably from the AB-*6b*, C-*6b*, AK-*6b* and T-*6b* phenotypes. The TE-*6b* genes did not induce tumours on *N. rustica* or on *K. tubiflora*. On media with 2% sucrose, TE-1-6b-L roots grew normally, unlike T-*6b* roots which strongly expand on such media (Tinland *et al.*, 1989, 1990, 1992; Grémillon *et al.*, 2004). T-*6b* root expansion is probably due to increased uptake and the accumulation of large amounts of sucrose (Clément *et al.*, 2006, 2007; Pasternak *et al.*, 2017).

In tobacco, TE-*6b* genes did not induce enations, catacorollas or ectopic vascular strands. Instead, TE-1-6b-R and TE-2-6b leaves showed strongly developed tertiary and higher-order veins, wrinkling and leaflets growing from petiole wings. Wrinkling is also typical for plants regenerated from hairy roots, although these do not contain *6b*

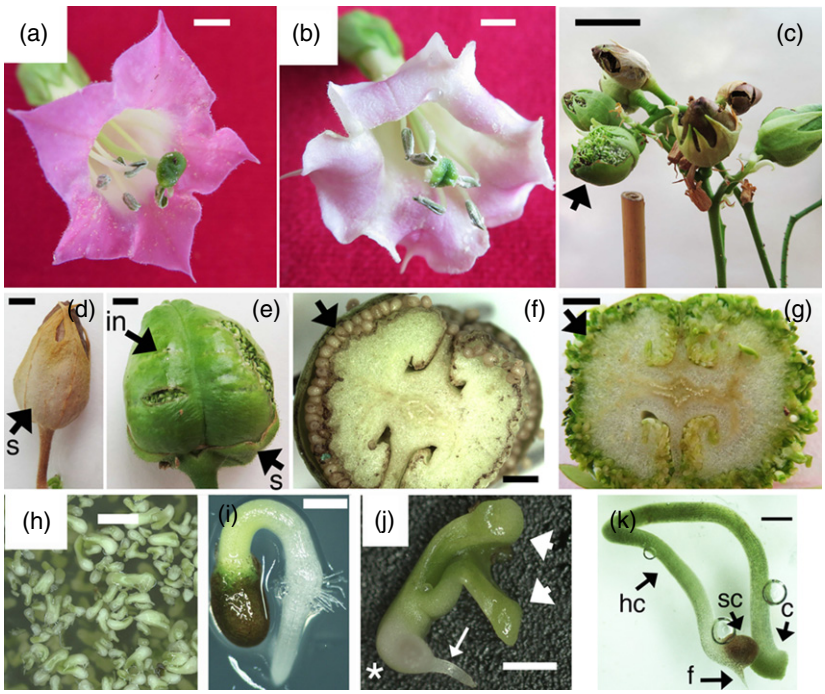

Figure 7. Flowers and seeds of TE-1-6b-L plants.
(a) Normal tobacco flower. (b) TE-1-6b-L flower. (c) Seed capsules on a TE-1-6b-L plant at various stages of development. Arrow: capsule with germinating embryos. (d) Normal tobacco seed capsule (s, sepal). (e) TE-1-6b-L seed capsule. This capsule is swollen, and partially split open with green embryos inside (in, integument; s, sepal, partly removed). (f) Normal tobacco seed capsule, transverse cut. Arrow: immature seeds at the periphery. (g) TE-1-6b-L seed capsule, transverse cut. Arrow: green, embryo-like structures. (h) Isolated TE-1-6b-L embryos released from a young capsule and placed in water. (i) Germinating tobacco seed; the radicle emerges first. (j) Abnormal germinated embryo from a TE-1-6b-L capsule (large arrows, cotyledons; small arrow, funiculus; asterisk, rootless base of embryo). (k) Embryo with a long hypocotyl (hc) and a short apex (c) (f, funiculus; sc, seed coat). The apical part has emerged first.
Scale bars: (a), (b), (d), (e) 5 mm; (c) 3 cm; (f), (g) 2 mm; (h) 1 mm; (i)–(k) 0.5 mm.

genes. TE-1-6b-L plants showed similar changes, but in addition formed large amounts of glandular trichomes, leaflets on the leaf edges, leaf spikes, split leaf tips, ectopic shoot primordia and modified cotyledons. Their flowers were modified, and most strikingly they became viviparous. Prematurely germinating embryos formed aberrant plantlets in the seed capsules, a phenomenon which is controlled by the parental plant. TE-1-6b-L plants could become an interesting tool for studying embryo growth. TE-1-6b-L embryos germinate at a very early stage and can be isolated in large numbers from sterilized capsules. This will allow studies on the external factors that influence their growth and development. The stalked spherical structures reported here seem to be an unusual form of leaf primordium. No roots were found on embryos, even after prolonged growth *in vitro*. However, when grown from seeds, seedlings with highly abnormal cotyledons formed normal roots. At the base of the embryos, a large spherical mass of colourless filaments appeared: these seem to correspond to root hairs.

Vivipary has been linked to deficiency of abscisic acid (ABA), to ABA insensitivity or to low osmotic potential (Durantini *et al.*, 2008; Wang *et al.*, 2016). It will be worthwhile to investigate these factors in TE-1-6b-L plants.

The induction and development of the apical stem meristems and of the basal filaments, and the formation of veins and dentate leaves are particularly interesting for basic studies on plant morphogenesis. This will necessitate further anatomical and physiological analysis, preferably by using an inducible TE-1-*6b-L* gene construct.

The dentate leaves of TE-1-6b-L tobacco plants resemble certain mutants of the miR319-jaw-TCP module in Arabidopsis, *Antirrhinum majus* and tomato. Overexpression of *MIR319A* in Arabidopsis (*jaw-D* mutants) or the combined mutation of the five miR319 target genes *TCP2*, *TCP3*, *TCP4*, *TCP10* and *TCP24*, leads to folded and crinkled leaves with outgrowth of the leaf margins. This is due to the abnormal maintenance of the leaf marginal meristem in such mutants (Nath *et al.*, 2003; Palatnik *et al.*, 2003; Alvarez *et al.*, 2016; Bresso *et al.*, 2017). Leaf vasculature is also increased (Alvarez *et al.*, 2016; Bresso *et al.*, 2017) as in the TE-6b plants. Possibly, TE-*6b* genes affect the miR319/jaw-TCP module; this will require introduction and further study of the TE-*6b* genes in Arabidopsis. The growth stimulation of higher-order leaf veins in TE-6b plants leads to a dense, regularly branched pattern, up to the leaf rim. This could rigidify the leaf before it is fully expanded, leading to wrinkling and outgrowth of spikes. Most likely, the strongly modified venation patterns redirect sucrose, amino acids and hormones to abnormal sites, affecting source–sink relations. This could explain the slow overall growth, the smaller dark-green leaves of these plants and the large number of glandular trichomes. Vivipary could be due to modifications in the vascular strands of the funiculi, leading to increased transport of nutrients from the placental tissues into the growing embryos. Non-synchronous development of the vascular connections between the placenta and the embryos could lead to patches with normal seeds or germinating embryos within the same capsule. The striking effects of TE-1-*6b-L* on

embryo development may be used to modify seed filling in crop plants.

The complex enation syndrome of T-*6b*, AB-*6b* and AKE-*6b* on the one hand, and the phenotypes of the TE-*6b* genes described here, are remarkably different. The role of the various amino acid residues in 6B proteins may be explored using hybrid proteins and proteins with mutations in specific residues. This may allow the identification of determinants for the two 6B phenotypes. No fewer than 46 TE-6B amino acid residues (out of 216) are unique compared with the corresponding residues of 18 other 6B sequences. Five residues are fully conserved in the non-TE-6B proteins, but are different in the TE-6B proteins (positions from TE-2-6B, TE-2-6B residues in brackets): Q26 (A), R28 (K), D51 (N), Y54 (C) and I147 (N). These could play a role in generating the observed phenotypic differences. The highly variable C-terminal acidic region of 6B proteins (Figure S5) is of special interest. The acidic region of AK-6B is essential for nuclear localization, transactivation and induction of hormone independence in tobacco (Kitakura *et al.*, 2002). The size and composition of this region could have important effects on the morphogenetic activities, which remain to be explored. Amino acid residues found earlier to be essential for T-6B activity (T92, P95, P96, F130 and A132; Helfer *et al.*, 2002) are conserved in the TE-6B proteins.

T-*6b* plants show enhanced sucrose uptake. If this also occurs in TE-6b plants it is most likely restricted to leaf tissues, since root growth is not modified. A link possibly exists between the initiation of ectopic veins, as in AB-*6b*, AK-*6b*, and T-*6b* plants (Helfer *et al.*, 2003; Terakura *et al.*, 2006; Chen and Otten, 2016), and stimulation of vein growth induced by TE-*6b* (this work). Interestingly, tobacco plants with strong AK-*6b* expression develop enations on the cotyledons, whereas plants with weak expression lack enations but show increased growth of cotyledon veins (Kakiuchi *et al.*, 2007). A possible role for TE-*6b* in auxin synthesis or transport, as proposed for AK-*6b* (Kakiuchi *et al.*, 2006, 2007), merits more detailed analysis.

It will also be important to study the cellular localization of the TE-6B proteins and their spatio-temporal expression patterns in *N. otophora*. Expression of the TE-*6b* genes under their own promoters will provide clues to their respective roles in *N. otophora*. The expression of at least two TE-*6b* genes in *N. otophora* leaves, the strong TE-*6b*-induced phenotype in leaves of the closely related *N. tabacum* and the strong venation of *N. otophora* leaves all point to a role for these genes in the growth and development of the natural transformant *N. otophora.*

The other *N. otophora plast* genes also require functional analysis. The three TE-*orf14* genes are especially interesting, since they are different from the usual *orf14* genes. The *orf14* genes are generally considered to have little or no biological activity (Aoki *et al.*, 1994; Aoki and Syono, 1999a), but their conservation in *N. otophora* and in other *Nicotiana* species (Chen *et al.*, 2014) suggests a biological function. The TE-*orf13* and TE-*rolC* genes may also have morphogenetic activity, like *orf13* and *rolC* from *N. glauca* and *N. tabacum*. We have proposed that the highly morphogenic T-DNA *plast* genes could have led to speciation, an essential step in the establishment of natural transformants (Chen and Otten, 2017). Various unusual *N. otophora* features, such as flower morphology, nectar production and pollinator choice, and the switch from day flowering to night flowering, could have contributed to speciation. In addition, *N. otophora* shows prominent tertiary and higher-order leaf veins and leaf wrinkling, similar to TE-6b tobacco plants. It is tempting to speculate that the TC and TE genes could have played a role in the evolutionary origin of these features.

RNA interference or CRISPR-Cas9 studies will provide a better insight in the role of individual *N. otophora* cT-DNA genes and gene combinations. The combined effect of the TC and TE genes can be studied by their transfer to other plant species or by complete removal from the *N. otophora* genome.

EXPERIMENTAL PROCEDURES

Mapping of cT-DNA regions and border analysis

For cT-DNA mapping, the *N. otophora* whole-genome sequence (WGS) AWOL series of contigs (Sierro *et al.*, 2014) was used. Individual contigs were extended by searching the main read collection (SRX335465) for overlapping sequences. Contigs were further assembled using *A. tumefaciens, A. rhizogenes and A. vitis* T-DNA regions as models. Mapping of reads on these sequences allowed separation of the different repeats. Subsequently, repetition-specific sequences were identified and used to amplify and sequence specific repeats and transition zones between these repeats and adjacent single-copy regions. Assembled regions were checked by re-mapping the individual reads on the assembled sequences. The TC-1, TC-2, TE-1 and TE-2 sequences are available upon request. A search for T-DNA border sequences was done with the YGRCAGGATATATNNNNNKGTMAWN border consensus (Barrell *et al.*, 2007).

RT-qPCR analysis

The RT-qPCR analysis of leaves of individual plants was done according to Chen *et al.* (2014). Young leaves (5 cm long) of soil-grown plants were used. For the three TE-*6b* genes, RT-qPCR forward (aagaggcggcagatgatg), and reverse (cgtagtccccgatttggtatt) primers were used. Expression of the internal standard gene *EF-2* was measured with forward (ctgaaccagaagcgtggaca) and reverse (ccagatgtagcagccctcaag) primers.

Transcriptional analysis

Seeds of *N. otophora* (TW95, PI235553) were sterilized with bleach (13% active chlorine) and fuming HCl and sown on rooting medium (Murashige and Skoog medium, Duchefa M0255, https://www.duchefa-biochemie.com/) and cultivated in a phytotronic chamber (Strader, strader.fr) at 24°C in 16-h light/8-h dark for 2 weeks until the plantlets were fully developed. RNA was extracted from roots and leaves using the RNeasy Plant Mini kit from Qiagen, according to the manufacturer's instructions. RNA

sequencing libraries were prepared using the TruSeq Stranded Total RNA Library Prep Kit from Illumina, and sequenced on an Illumina HiSeq2500 in rapid mode. RNA-seq data were mapped to the reference TC-1, TC-2, TE-1 and TE-2 sequences using hisat2 and filtered to keep only perfectly and uniquely mapped unspliced reads. Twenty-eight per cent of the unspliced leaf reads and 33% of the unspliced root reads mapping to the T-DNA were removed when filtering for perfect and unique mapping.

Cloning of TE-*6b* genes

Cloning of TE-*6b* genes was carried out in two steps. First, each region carrying a TE-*6b* gene was amplified separately by PCR using region-specific primers. The three TE-*6b* ORFs were then amplified with appropriate primers and inserted in the pCK vector carrying a 2 × 35S promoter and a termination site (Carrington *et al.*, 1991). Constructs were sequenced and TE-*6b* genes were introduced into the binary vector pBI121.2 (Jefferson *et al.*, 1987). Binary vector derivatives were introduced into LBA4404 (Hoekema *et al.*, 1983) for transformation.

Protein and DNA alignments

DNA sequence alignments for *orf3n-orf8* and protein alignments for 6B sequences were done with Multialign (Corpet, 1988). Plast protein sequences were aligned with MUSCLE (Edgar, 2004).

Inoculation and transformation

Stems of *N. rustica*, *K. tubiflora* and *Kalanchoe daigremontiana* were inoculated as reported (Helfer *et al.*, 2002). For transformation, Samsun nn leaf discs were incubated with the various LBA4404 derivatives for 2 days, washed with 350 mg l^{-1} Claforan and placed on M0255 medium (Duchefa) with 100 mg l^{-1} kanamycin, 350 mg l^{-1} Claforan, NAA (0.05 mg l^{-1}) and BAP (2 mg l^{-1}). Shoots were transferred to M0255 with 100 mg l^{-1} kanamycin and 350 mg l^{-1} Claforan without hormones. After rooting, plantlets were transferred to soil. F_1 seedlings were grown on M0222.

Embryo cultures

TE-1-6b-L embryo cultures were started from TE-1-6b-L line S39 capsules, early after pollination. After removal of the sepals, capsules were rinsed with 70% ethanol, then sterilized for 20 min with bleach and 0.01% Tween-20, and washed three times (20 min each time) with distilled water. Embryos were placed on M0222 medium with 1% sucrose.

ACKNOWLEDGEMENTS

KC was supported by doctoral grant 2011679003 from the Chinese Scholarship Council. We thank the IBMP gardeners for providing plants, IPK Gatersleben, and Jessica Nifong of the US Nicotiana Germplasm Collection for *N. otophora* seeds.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Leaf venation pattern of *Nicotiana otophora*.

Figure S2. Alignment of TC *orf3n-orf8* regions.

Figure S3. Phylogenetic tree of Plast proteins.

Figure S4. Transcription patterns of TC-1, TC-2, TE-1 and TE-2.

Figure S5. Alignment of 6B sequences.

Figure S6. Morphological changes of TE-1-6b-L leaves.

Figure S7. Phenotypes of TE-1-6b-L lines, TE-1-*6b-L* expression, and root growth.

Figure S8. Modified seed capsules of TE-1-6b-L plants.

Figure S9. Embryos from TE-1-6b-L plants.

Figure S10. TE-1-6b-L embryos *in vitro*.

REFERENCES

Alvarez, J. P., Furumizu, C., Efroni, I., Eshed, Y. and Bowman, J. L. (2016) Active suppression of a leaf meristem orchestrates determinate leaf growth. *eLife*, **5**, e15023.

Aoki, S. (2004) Resurrection of an ancestral gene: functional and evolutionary analyses of the Ng*rol* genes transferred from *Agrobacterium* to *Nicotiana*. *J. Plant. Res.* **117**, 329–337.

Aoki, S. and Syono, K. (1999a) Synergistic function of *rolB*, *rolC*, *ORF13,* and *ORF14* of TL-DNA of *Agrobacterium rhizogenes* in hairy root induction in *Nicotiana tabacum*. *Plant Cell Physiol.* **40**, 252–256.

Aoki, S. and Syono, K. (1999b) Function of Ng*rol* genes in the evolution of *Nicotiana glauca*: conservation of the function of NgORF13 and NgORF14 after ancient infection by an *Agrobacterium rhizogenes*-like ancestor. *Plant Cell Physiol.* **40**, 222–230.

Aoki, S. and Syono, K. (1999c) Horizontal gene transfer and mutation: Ng*rol* genes in the genome of *Nicotiana glauca*. *Proc. Natl Acad. Sci. USA*, **96**, 13229–13234.

Aoki, S. and Syono, K. (2000) The roles of Ri*rol* and Ng*rol* genes in hairy root induction in *Nicotiana debneyi*. *Plant Sci.* **159**, 183–189.

Aoki, S., Kawaoka, A., Sekine, M., Ichikawa, T., Fujita, T., Shinmyo, A. and Syono, K. (1994) Sequence of the cellular T-DNA in the untransformed genome of *Nicotiana glauca* that is homologous to *ORFs 13* and *14* of the Ri plasmid and analysis of its expression in genetic tumors of *N. glauca* x *N. langsdorffii*. *Mol. Gen. Genet.* **243**, 706–710.

Barrell, P.J., Jacobs, J.M.E., Baldwin, S. and Conner, A.J. (2007) Intragenic vectors of gene transfer without foreign DNA. *Euphytica*, **154**, 341–353.

Bresso, E. G., Chorostecki, U., Rodriguez, R. E., Palatnik, J. F. and Schommer, C. (2017) Spatial control of gene expression by miR319-regulated TCP transcription factors in leaf development. *Plant Physiol.* **176**, 1694–1708. Epub ahead of print.

Canaday, J., Gérard, J.C., Crouzet, P. and Otten, L. (1992) Organization and functional analysis of three T-DNAs from the vitopine Ti plasmid pTiS4. *Mol. Gen. Genet.* **235**, 292–303.

Carrington, J.C., Freed, D.D. and Leinicke, A.J. (1991) Bipartite signal sequence mediates nuclear translocation of the plant potyviral NIa protein. *Plant Cell*, **3**, 953–962.

Chen, K. (2016) Sequencing and functional analysis of cT-DNAs in *Nicotiana*. Thesis of University of Strasbourg.

Chen, K. and Otten, L. (2016) Morphological analysis of the *6b* oncogene-induced enation syndrome. *Planta*, **243**, 131–148.

Chen, K. and Otten, L. (2017) Natural *Agrobacterium* transformants: recent results and some theoretical considerations. *Front. Plant Sci.* **8**, 1600.

Chen, K., Dorlhac de Borne, F., Szegedi, E. and Otten, L. (2014) Deep sequencing of the ancestral tobacco species *Nicotiana tomentosiformis* reveals multiple T-DNA inserts and a complex evolutionary history of natural transformation in the genus *Nicotiana*. *Plant J.* **80**, 669–682.

Chen, K., Dorlhac de Borne, F., Julio, E., Obszynski, J., Pale, P. and Otten, L. (2016) Root-specific expression of opine genes and opine accumulation in some cultivars of the naturally occurring GMO *Nicotiana tabacum*. *Plant J.* **87**, 258–269.

Clément, B., Pollmann, S., Weiler, E., Urbanczyk-Wochniak, E. and Otten, L. (2006) The *Agrobacterium vitis* T-6b oncoprotein induces auxin-independent cell expansion in tobacco. *Plant J.* **45**, 1017–1027.

Clément, B., Perot, J., Geoffroy, P., Legrand, M., Zon, J. and Otten, L. (2007) Abnormal accumulation of sugars and phenolics in tobacco roots expressing the *Agrobacterium* T-*6b* oncogene and the role of these compounds in *6b*-induced growth. *Mol. Plant-Microbe Interact.* **20**, 53–62.

Corpet, F. (1988) Multiple sequence alignment with hierarchical clustering. *Nucl. Acids Res.* **16**, 10881–10890.

Durantini, D., Guiliani, A., Malgioglio, A., Pilu, R., Tuberosa, R., Sanguineti, C. and Gavazzi, G. (2008) Vivipary as a tool to analyze late embryonic events in maize. *Heredity*, **101**, 465–470.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* **32**, 1792–1797.

Fründt, C., Meyer, A.D., Ichikawa, T. and Meins, F. (1998a) Evidence for the ancient transfer of Ri plasmid T-DNA genes between bacteria and plants. In *Horizontal Gene Transfer*. (Syvanen, M. and Kado, C. I., eds). London: Chapman and Hall, pp. 94–106.

Fründt, C., Meyer, A.D., Ichikawa, T. and Meins, F. (1998b) A tobacco homologue of the Ri-plasmid *orf13* gene causes cell proliferation in carrot root discs. *Mol. Gen. Genet.* **259**, 559–568.

Furner, I.J., Huffman, G.A., Amasino, R.M., Garfinkel, D.J., Gordon, M.P. and Nester, E.W. (1986) An *Agrobacterium* transformation in the evolution of the genus *Nicotiana*. *Nature*, **319**, 422–427.

Gális, I., Simek, P., Van Onckelen, H.A., Kakiuchi, Y. and Wabiko, H. (2002) Resistance of transgenic tobacco seedlings expressing the *Agrobacterium tumefaciens* C58-*6b* gene to growth-inhibitory levels of cytokinin is associated with elevated IAA levels and activation of phenylpropanoid metabolism. *Plant Cell Physiol.* **43**, 939–950.

Gális, I., Kakiuchi, Y., Simek, P. and Wabiko, H. (2004) *Agrobacterium tumefaciens* AK-*6b* gene modulates phenolic compound metabolism in tobacco. *Phytochem.* **65**, 169–179.

Goodspeed, T.H. (1954) *The Genus Nicotiana*. Waltham, Massachusetts, USA: Chronica Botanica.

Grémillon, L., Helfer, A., Clément, B. and Otten, L. (2004) New plant growth-modifying properties of the *Agrobacterium* T-*6b* oncogene revealed by the use of a dexamethasone-inducible promoter. *Plant J.* **37**, 218–228.

Grisebach, A. (1879) Symbolae ad Floram Argentinam. Zweite Bearbeitung argentinischer Pflanzen. *Abhandlungen der Königlichen Gesellschaft der Wissenschaften zu Göttingen* **24**, 243.

Hansen, G., Larribe, M., Vaubert, D., Tempé, J., Biermann, B.J., Montoya, M.D., Chilton, M.D. and Brevet, J. (1991) *Agrobacterium rhizogenes* pRi8196 T-DNA: mapping and DNA sequence of functions involved in mannopine synthesis and hairy root differentiation. *Proc. Natl Acad. Sci. USA*, **88**, 7763–7767.

Helfer, A., Pien, S. and Otten, L. (2002) Functional diversity and mutational analysis of *Agrobacterium 6B* oncoproteins. *Mol. Genet. Genom.* **267**, 577–586.

Helfer, A., Clément, B., Michler, P. and Otten, L. (2003) The *Agrobacterium* oncogene AB-*6b* causes a graft-transmissible enation syndrome in tobacco. *Plant Mol. Biol.* **52**, 483–493.

Hoekema, A., Hirsch, P.R., Hooykaas, P.J.J. and Schilperoort, R.A. (1983) A binary plant vector strategy based on separation of vir- and T-region of the *Agrobacterium tumefaciens* Ti-plasmid. *Nature*, **303**, 179–180.

Hooykaas, P.J.J., den Dulk-Ras, H. and Schilperoort, R.A. (1988) The *Agrobacterium tumefaciens* T-DNA *6b* is an oncogene. *Plant Mol. Biol.* **11**, 791–794.

Ishibashi, N., Kitakura, S., Terakura, S., Machida, C. and Machida, Y. (2014) Protein encoded by oncogene *6b* from *Agrobacterium tumefaciens* has a reprogramming potential and histone chaperone-like activity. *Front. Plant Sci.* **5**, 1–7.

Ito, M. and Machida, Y. (2015) Reprogramming of plant cells induced by 6b oncoproteins from the plant pathogen *Agrobacterium*. *J. Plant. Res.* **128**, 423–435.

Jefferson, R.A., Kavanagh, T.A. and Bevan, M. (1987) GUS fusions: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* **6**, 3901–3907.

Kakiuchi, Y., Gális, I., Tamogami, S. and Wabiko, H. (2006) Reduction of polar auxin transport in tobacco by the tumorigenic *Agrobacterium tumefaciens* AK-*6b* gene. *Planta*, **223**, 237–247.

Kakiuchi, Y., Takahashi, S. and Wabiko, H. (2007) Modulation of the venation pattern of cotyledons of transgenic tobacco for the tumorigenic *6b* gene of *Agrobacterium tumefaciens* AKE10. *J. Plant. Res.* **120**, 259–268.

Kitakura, S., Fujita, T., Ueno, Y., Terakura, S., Wabiko, H. and Machida, Y. (2002) The protein encoded by oncogene *6b* from *Agrobacterium tumefaciens* interacts with a nuclear protein of tobacco. *Plant Cell*, **14**, 451–463.

Kitakura, S., Terakura, S., Yoshioka, Y., Machida, C. and Machida, Y. (2008) Interaction between *Agrobacterium tumefaciens* oncoprotein 6b and a tobacco nucleolar protein that is homologous to TNP1 encoded by a transposable element of *Antirrhinum majus*. *J. Plant. Res.* **121**, 425–433.

Knapp, S., Chase, M.W. and Clarkson, J.J. (2004) Nomenclatural changes and a new sectional classification in *Nicotiana* (Solanaceae). *Taxon*, **53**, 73–82.

Kyndt, T., Quispe, D., Zhai, H., Jarret, R., Ghislain, M., Liu, Q., Gheysen, G. and Kreuze, J.F. (2015) The genome of cultivated sweet potato contains *Agrobacterium* T-DNAs with expressed genes: an example of a naturally transgenic food crop. *Proc. Natl Acad. Sci. USA*, **112**, 5844–5849.

Lemcke, K. and Schmülling, T. (1998) Gain of function assays identify non-*rol* genes from *Agrobacterium rhizogenes* TL-DNA that alter plant morphogenesis or hormone sensitivity. *Plant J.* **15**, 423–433.

Levesque, H., Delepelaire, P., Rouzé, P., Slightom, J. and Tepfer, D. (1988) Common evolutionary origin of the central portion of the Ri TL-DNA of *Agrobacterium rhizogenes* and the Ti T-DNAs of *Agrobacterium tumefaciens*. *Plant Mol. Biol.* **11**, 731–744.

Lim, K.Y., Kovarik, A., Matyasek, R., Chase, M.W., Clarkson, J.J., Grandbastien, M.A. and Leitch, A.R. (2007) Sequence of events leading to near-complete genome turnover in allopolyploid Nicotiana within five million years. *New Phytol.* **175**, 756–763.

Long, N., Ren, X., Xiang, Z., Wan, W. and Dong, Y. (2016) Sequencing and characterization of leaf transcriptomes of six diploid *Nicotiana* species. *J. Biol. Res. Thessaloniki*, **23**, 1–12.

Matveeva, T.V. and Lutova, L.A. (2014) Horizontal gene transfer from *Agrobacterium* to plants. *Front. Plant Sci.* **8**, 326.

Matveeva, T.V., Bogomaz, D.I., Pavlova, O.A., Nester, E.W. and Lutova, L.A. (2012) Horizontal gene transfer from genus *Agrobacterium* to the plant *Linaria* in nature. *Mol. Plant-Microbe Interact.* **25**, 1542–1551.

Meyer, A.D., Ichikawa, T. and Meins, F. (1995) Horizontal gene transfer: regulated expression of tobacco homologue of the *Agrobacterium rhizogenes rolC* gene. *Mol. Gen. Genet.* **249**, 265–273.

Mohajjel-Shoja, H. (2010) Contribution to the study of the *Agrobacterium rhizogenes plast* genes, *rolB* and *rolC*, and their homologs in *Nicotiana tabacum*. Thesis of University of Strasbourg.

Mohajjel-Shoja, H., Clément, B., Perot, J., Alioua, M. and Otten, L. (2011) Biological activity of the *Agrobacterium rhizogenes*-derived *trolC* gene of *Nicotiana tabacum* and its functional relationship to other *plast* genes. *Mol. Plant-Microbe Interact.* **24**, 44–53.

Nath, U., Crawford, B.C.W., Carpenter, R. and Coen, E. (2003) Genetic control of surface curvature. *Science*, **299**, 1404–1407.

Nattero, J., Moré, M., Sersic, A.N. and Cocucci, A.A. (2003) Possible tobacco progenitors share long-tongued hawkmoths as pollen vectors. *Plant Syst. Evol.* **241**, 47–54.

Palatnik, J.F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J.C. and Weigel, D. (2003) Control of leaf morphogenesis by microRNAs. *Nature*, **425**, 257–263.

Pasternak, T., Haser, T., Falk, T., Ronneberger, O., Palme, K. and Otten, L. (2017) A 3D digital atlas of the *Nicotiana tabacum* root tip and its use to investigate changes in the root apical meristem induced by the *Agrobacterium 6b* oncogene. *Plant J.* **92**, 31–42.

Reed, S.M. and Schneider, J. (1992) Evaluation of *Nicotiana otophora* as a source of resistance to *Meloidogyne incognita* Race 4 for tobacco. *J. Nematol.* **24**, 253–256.

Sierro, N., Battey, J.N., Ouadi, S., Bakaher, N., Bovet, L., Willig, A., Goepfert, S., Peitsch, M.C. and Ivanov, N.V. (2014) The tobacco genome sequence and its comparison with those of tomato and potato. *Nat. Commun.* **5**, 3833.

Slightom, J.L., Durand-Tardif, M., Jouanin, L. and Tepfer, D. (1986) Nucleotide sequence analysis of TL-DNA of *Agrobacterium rhizogenes* agropine type plasmid. Identification of open reading frames. *J. Biol. Chem.* **261**, 108–121.

Takahashi, S., Sato, R., Takahashi, M., Hashiba, N., Ogawa, A., Toyofuku, K., Sawata, T., Ohsawa, Y., Ueda, K. and Wabiko, H. (2013) Ectopic localization of auxin and cytokinin in tobacco seedlings by the plant-oncogenic AK-*6b* gene of *Agrobacterium tumefaciens* AKE10. *Planta*, **238**, 753–770.

Terakura, S., Kitakura, S., Ishikawa, M., Ueno, Y., Fujita, T., Machida, C., Wabiko, H. and Machida, Y. (2006) Oncogene *6b* from *Agrobacterium tumefaciens* induces abaxial cell division at late stages of leaf

development and modifies vascular development in petioles. *Plant Cell Physiol.* **47**, 664–672.

Terakura, S., Ueno, Y., Tagami, H. *et al.* (2007) An oncoprotein from the plant pathogen *Agrobacterium* has histone chaperone-like activity. *Plant Cell*, **19**, 2855–2865.

Tiedge, K. and Lohaus, G. (2017) Nectar sugars and amino acids in day- and night-flowering *Nicotiana* species are more strongly shaped by pollinator's preferences than organic acids and inorganic ions. *PLoS ONE*, **12** (5), e01176865.

Tinland, B., Huss, B., Paulus, F., Bonnard, G. and Otten, L. (1989) *Agrobacterium tumefaciens 6b* genes are strain-specific and affect the activity of auxin as well as cytokinin genes. *Mol. Gen. Genet.* **219**, 217–224.

Tinland, B., Rohfritsch, O., Michler, P. and Otten, L. (1990) *Agrobacterium tumefaciens* T-DNA gene *6b* stimulates rol-induced root formation, permits growth at high auxin concentrations and increases root size. *Mol. Gen. Genet.* **223**, 1–10.

Tinland, B., Fournier, P., Heckel, T. and Otten, L. (1992) Expression of a chimaeric heat-shock-inducible *Agrobacterium 6b* oncogene in *Nicotiana rustica*. *Plant Mol. Biol.* **18**, 921–930.

Wabiko, H. and Minemura, M. (1996) Exogenous phytohormone-independent growth and regeneration of tobacco plants transgenic for the *6b* gene of *Agrobacterium tumefaciens* AKE10. *Plant Physiol.* **112**, 939–951.

Wang, M., Soyano, T., Machida, S., Yang, J.Y., Jung, C., Chua, N.H. and Yuan, Y.A. (2011) Molecular insights into plant cell proliferation disturbance by *Agrobacterium* protein 6b. *Genes Dev.* **25**, 64–76.

Wang, X., Zhang, L., Xu, X., Qu, W., Li, J., Xu, X. and Wang, A. (2016) Seed development and viviparous germination in one accession of a tomato rin mutant. *Breed. Sci.* **66**, 372–380.

Wenke, T., Döbel, T., Sörensen, T.R., Junghans, H., Weisshaar, B. and Schmidt, T. (2011) Targeted identification of short interspersed nuclear element families show their widespread existence and extreme heterogeneity in plant genomes. *Plant Cell*, **23**, 3117–3128.

White, F.F., Garfinkel, D.J., Huffman, G.A., Gordon, M.P. and Nester, E.W. (1983) Sequence homologous to *Agrobacterium rhizogenes* T-DNA in the genomes of uninfected plants. *Nature*, **301**, 348–350.

Wolters, A.A., Uitdewilligen, J.G.A.M.L., Kloosterman, B.A., Hutten, R.C.B., Visser, R.G.F. and van Eck, H.J. (2010) Identification of alleles of carotenoid pathway genes important for zeaxanthin accumulation in potato tubers. *Plant Mol. Biol.* **73**, 659–671.